Robust Optimal Sensor Placement for Operational Modal Analysis Based on Maximum Expected Utility

Binbin Li

PhD Candidate, Civil and Environmental Engineering,

University of California, Berkeley, CA 94720. Phone: (510) 612-2490, E-mail: bbl@berkeley.edu

Armen Der Kiureghian^a

President, American University of Armenia, Yerevan, Armenia

Taisei Professor of Civil Engineering Emeritus

University of California, Berkeley, CA 94720

Phone: (510) 642-2469, Fax: (510) 643-5264, E-mail: adk@ce.berkeley.edu

Abstract: Optimal sensor placement is essentially a decision problem under uncertainty. The maximum expected utility theory and a Bayesian linear model are used in this paper for robust sensor placement aimed at operational modal identification. To avoid nonlinear relations between modal parameters and measured responses, we choose to optimize the sensor locations relative to identifying modal responses. Since the modal responses contain all the information necessary to identify the modal parameters, the optimal sensor locations for modal response estimation provide at least a suboptimal solution for identification of modal parameters. First, a probabilistic model for sensor placement considering model uncertainty, load uncertainty and measurement error is proposed. The maximum expected utility theory is then applied with this model by considering utility functions based on three principles: Shannon information, quadratic loss, and K-L divergence. In addition, the prior covariance of modal responses under band-limited white-noise excitation is derived and the nearest Kronecker product approximation is employed to accelerate evaluation of the utility function. As demonstration and validation examples, sensor placements in a 16-degrees-of-freedom shear-type building and in Guangzhou TV Tower under ground motion and wind load are considered. Placements of individual displacement meter, velocimeter, accelerometer and placement of mixed sensors are illustrated.

Keywords: Optimal Sensor placement, Maximum Expected Utility, Bayesian Inference, Operational Modal Identification, Prior Information, Band-limited White Noise

1. Introduction

Modal parameter identification is essential for system identification, finite element model updating and damage identification [1]. In order to identify modal parameters (natural frequencies, damping ratios and mode shapes) from measured structural responses, a sensor system needs to be designed in advance. Although with the advent of sensor technology the cost of sensors is rapidly decreasing, the affordable number of sensors is still an issue in long-span bridges and super-tall buildings. When a limited number of sensors is used, inevitably uncertainties arise in the identified parameters. Many researches [2, 3] have shown that optimal sensor

^aCorresponding Author

placement (OSP) is of crucial importance in order to improve the accuracy and precision of modal parameter identification.

OSP can be formally defined as an optimization problem consisting of performance criteria and computational issues. Among various formulations of performance criteria, information theory based approaches have attracted most attention. Kammer [2] developed the effective independence (EI) method for modal identification, which tends to maximize the determinant of the Fisher information matrix (FIM). Udwadia [3] proposed the Fisher information criterion for OSP in parameter identification, in which the optimal configuration corresponds to that maximizing the trace of the FIM. Fisher information was also introduced by Borguet & Léonard [4] in the field of engine health monitoring, where the weighted sum of the condition number, trace and determinant of the FIM was selected as the performance criterion. Papadimitriou [5] introduced the concept of information entropy for the purpose of minimizing the uncertainty in the model parameter estimation; the effect of prediction error correlation was further examined recently by Papadimitriou & Lombaert [6]. In addition, Trendafilova, et al. [7] employed mutual entropy to select sensor locations to produce independent measurements. Li [8] derived the expected Kullback-Leibler divergence criterion to deploy sensors, which turns out to be identical to the weighted sum of the determinant and trace of the FIM. A comprehensive discussion and comparison of information-theory based OSP can be found in Li [9]. Aside from these information-theory based approaches, many other methods are proposed based on modal kinetic energy [10, 11, 12], model reduction [13, 14] and observability [15, 16]. The literature on performance criteria for OSP is so extensive that we cannot list all references. The interested reader may refer to Li [17] for additional references.

Structural responses are distinct under different input loads; thus, the optimal sensor configuration generally depends on the load case. Li [18] proposed a load-dependent sensor placement method considering both the load and structural response, and showed improved identification performance. Brehm [19] determined the optimal locations of a reference sensor under white-noise excitation and multiple impulse excitations. Furthermore, the nominal model of the structure is invariably biased, so that a nominal model-based approach cannot provide a robust design. Vinot [20] introduced a test planning procedure based on info-gap decision theory to optimize the worst possible performance for all realizations of model parameters. Castro-Triguero [21] examined the influence of model parameters on OSP using Monte Carlo simulation. The Bayesian approach has also been employed to tackle these problems. Heredia-Zavoni [22] used the expected Bayesian loss function to deploy sensors for parameter identification under seismic load considering uncertainties in stiffness. Yuan et al. [23] proposed an information entropy-based OSP methodology for modal identification considering an uncertain excitation, wherein the model parameter uncertainties were investigated by Monte Carlo sampling. Flynn [24] placed sensors in order to minimize the expected total presence of either type I or type II errors during the damage detection process using guided ultrasonic waves.

Even though great progress has been made in optimal sensor placement, many aspects of the problem remain unresolved. For example, mode-shape-based sensor placement approaches cannot incorporate prior information about natural frequencies and damping ratios. Secondly, in the formulation of FIM, independence of measured responses is assumed. This assumption obviously is not appropriate because different responses of a structure to the same excitation are naturally correlated. Thirdly, an OSP design based on a nominal model of the structure can be over-optimistic or even misleading in certain cases. Fourthly, most methods consider deployment of displacement meter, velocimeter and accelerometer to have the same effect for modal identi-

fication, but intuition suggests that this may not be true. Furthermore, although Bayesian approaches have been used [22, 23, 24], they have only employed diffuse priors, which essentially degenerates the problem into the traditional FIM-based approach. In this work, we attempt to address all the above drawbacks by utilizing the maximum expected utility theory [25], which is a perfect combination of decision theory and Bayesian statistics. Decision theory provides the mathematical foundation for selection of the optimal design, while the Bayesian approach provides a coherent framework where prior information and uncertainties regarding unknown quantities can be combined to provide the needed probabilistic information.

In this paper, we focus on OSP for operational modal analysis (OMA) [26], which due to its economy and efficiency has become the primary modal testing method in civil engineering. In OMA, only structural responses under operating conditions are measured and the unmeasured force is modeled as a broad-band random process, in most cases as a zero-mean, band-limited white-noise (BWN) process. We first develop a probabilistic framework for sensor placement incorporating model, load and measurement uncertainties, and formulate various design objectives in terms of utility functions. Computational issues are then discussed, including calculation of the covariance matrix of modal responses under BWN and fast evaluation of the utility functions. To allow the use of a linear Bayesian model, we formulate the optimization problem in terms of modal responses, which contain all the information necessary to identify modal parameters. Optimal sensor placements of an example mass-spring system and of the Guangzhou TV Tower [27] are presented in order to demonstrate the proposed approach.

2. Probabilistic Model for Sensor Placement under Stochastic Excitation

The equation of motion of a discrete, linear, and time-invariant dynamical system with N_d degrees of freedom (DOFs) under stochastic external force and ground motion is described by

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{C}\dot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = \mathbf{P}\mathbf{f}(t) - \mathbf{M}\mathbf{l}\ddot{\mathbf{u}}_g(t); \mathbf{u}(0) = \mathbf{u}_0, \dot{\mathbf{u}}(0) = \dot{\mathbf{u}}_0 \quad (1)$$

where $\mathbf{M} \in \mathbb{R}^{N_d \times N_d}$, $\mathbf{C} \in \mathbb{R}^{N_d \times N_d}$ and $\mathbf{K} \in \mathbb{R}^{N_d \times N_d}$ are the mass, damping and stiffness matrices, respectively; $\mathbf{u}(t) \in \mathbb{R}^{N_d}$, $\dot{\mathbf{u}}(t) \in \mathbb{R}^{N_d}$ and $\ddot{\mathbf{u}}(t) \in \mathbb{R}^{N_d}$ are the nodal displacement, velocity and acceleration responses relative to the ground motion; $\mathbf{f}(t) \in \mathbb{R}^{N_f}$ is the external force vector; $\mathbf{P} \in \mathbb{R}^{N_d \times N_f}$ is the load coefficient matrix representing the spatial influence of the external force; $\ddot{\mathbf{u}}_g(t) \in \mathbb{R}^{N_g}$ is the ground acceleration vector, and $\mathbf{l} \in \mathbb{R}^{N_d \times N_g}$ is the corresponding influence matrix. $\mathbf{u}_0 \in \mathbb{R}^{N_d}$ and $\dot{\mathbf{u}}_0 \in \mathbb{R}^{N_d}$ are the initial relative displacement and velocity vectors. Note that we are considering both an external force and a base motion. This is because under operating conditions the structure is usually subject to the effects of wind, traffic, ground tremor and small magnitude earthquakes.

Assuming classical damping, the equations of motion are decoupled by introducing the transformation to modal coordinates $\mathbf{u}(t) = \Phi \mathbf{q}(t)$, yielding

$$\ddot{q}_i(t) + 2\xi_i\omega_i\dot{q}_i(t) + \omega_i^2q_i(t) = \gamma_i^e\mathbf{f}(t) + \gamma_i^g\ddot{\mathbf{u}}_g(t);$$

$$q_i(0) = \phi_i^T \mathbf{M} \mathbf{u}_0 / M_i, \dot{q}_i(0) = \phi_i^T \mathbf{M} \dot{\mathbf{u}}_0 / M_i, i = 1, \dots, N_d \quad (2)$$

where $\mathbf{q}(t) = [q_1(t), q_2(t), \dots, q_{N_d}(t)]^T \in \mathbb{R}^{N_d}$ is the modal displacement vector and $\Phi = [\phi_1, \phi_2, \dots, \phi_{N_d}] \in \mathbb{R}^{N_d \times N_d}$ is the mode shape matrix. Moreover, $M_i = \phi_i^T \mathbf{M} \phi_i$, $\omega_i = \sqrt{\phi_i^T \mathbf{K} \phi_i / M_i}$,

$\xi_i = \boldsymbol{\phi}_i^T \mathbf{C} \boldsymbol{\phi}_i / (2M_i \omega_i)$, $\gamma_i^e = \boldsymbol{\phi}_i^T \mathbf{P} / M_i$ and $\gamma_i^g = \boldsymbol{\phi}_i^T \mathbf{M} \mathbf{l} / M_i$ are the i th modal mass, natural frequency, damping ratio and participation factors for external force and base motion, respectively.

Let there be n sensors and $\mathbf{S} \in \mathbb{R}^{n \times N_d}$ denote the selection matrix for the placement of sensors so that $\ddot{\mathbf{u}}_{ms} = \mathbf{S} \ddot{\mathbf{u}}$ represents the measured accelerations. In monitoring applications, usually only a few well-excited modes are observable. Let m denote the number of such modes and $\mathbf{S}_q \in \mathbb{R}^{m \times N_d}$ denote the selection matrix so that $\mathbf{S}_q \mathbf{q}$ represents the coordinates and $\boldsymbol{\Phi} \mathbf{S}_q^T$ represents the eigenvectors of the selected modes. The product $\boldsymbol{\Phi} \mathbf{S}_q^T \mathbf{S}_q \ddot{\mathbf{q}}$ then gives the contributions of the selected modes to the acceleration response. Furthermore, responses are recorded in discrete time and measurement errors are ubiquitous. Letting T denote the sampling period, one can write for the measured accelerations at time $t = kT$, $k = 0, 1, \dots, N$,

$$\ddot{\mathbf{u}}_{ms}[k] = \mathbf{S}(\boldsymbol{\Phi} \mathbf{S}_q^T \mathbf{S}_q \ddot{\mathbf{q}}[k] + \mathbf{l} \ddot{\mathbf{u}}_g[k] + \mathbf{w}[k]) \quad (3)$$

where $\mathbf{w}[k]$ represents the joint contribution of the unobservable modes and the measurement noise. Hereafter, we assume $\mathbf{w}$ is a zero-mean band-limited Gaussian white-noise process with covariance matrix $\boldsymbol{\Sigma}$. Both $\mathbf{S}$ and $\mathbf{S}_q$ consist of only ones and zeroes so that $\mathbf{S}^T \mathbf{S}$ and $\mathbf{S}_q^T \mathbf{S}_q$ are diagonal matrices with unity in their i th diagonal element only if the i th DOF or i th mode is selected, and zero otherwise. A similar formulation can be used to represent measured displacement or velocity, or even a mixture of measurements. Here, we show the formulation for measured acceleration because it is the most common measured quantity used for modal identification. The measured responses are absolute quantities, and it is for that reason that the ground acceleration $\ddot{\mathbf{u}}_g$ appears in the equation. Because in most cases there will be an accelerometer to record the ground motion, we assume it to be known with some measurement uncertainty and reformulate Eqn. (3) to read

$$\ddot{\mathbf{u}}_{rs}[k] = \ddot{\mathbf{u}}_{ms}[k] - \mathbf{S} \mathbf{l} \ddot{\mathbf{u}}_g[k] = \mathbf{S} \boldsymbol{\Phi} \mathbf{S}_q^T \mathbf{S}_q \ddot{\mathbf{q}}[k] + \mathbf{S} \mathbf{w}[k] \quad (4)$$

where $\ddot{\mathbf{u}}_{rs}$ is now the “measured” relative acceleration. Note that the covariance matrix $\boldsymbol{\Sigma}$ of the measurement noise includes partial contribution from the error in measuring the ground acceleration in this new formulation. Another issue is that both $\ddot{\mathbf{q}}[k]$ and $\ddot{\mathbf{u}}_{rs}[k]$ are temporally correlated processes, hence we have to consider all N measurement instants simultaneously for identification of modal response. For the sake of convenience, we introduce the following notations:

$$\mathbf{Y}_{rs} = [\ddot{\mathbf{u}}_{rs}[1]; \dots; \ddot{\mathbf{u}}_{rs}[N]]; \quad \mathbf{X}_s = \mathbf{I}_N \otimes (\mathbf{S} \boldsymbol{\Phi} \mathbf{S}_q^T);$$

$$\boldsymbol{\theta}_s = [\mathbf{I}_N \otimes \mathbf{S}_q][\ddot{\mathbf{q}}[1]; \dots; \ddot{\mathbf{q}}[N]]; \quad \mathbf{W}_s = [\mathbf{I}_N \otimes \mathbf{S}][\mathbf{w}[1]; \dots; \mathbf{w}[N]];$$

in which $\mathbf{I}_N \in \mathbb{R}^{N \times N}$ is the identity matrix and $\otimes$ stands for the Kronecker product. Eqn. (4) then becomes

$$\mathbf{Y}_{rs} = \mathbf{X}_s \boldsymbol{\theta}_s + \mathbf{W}_s \quad (5)$$

in which $\boldsymbol{\theta}_s$ represents modal accelerations. With the input load modeled as a band-limited Gaussian white noise, $\boldsymbol{\theta}_s$ follows a multivariate normal distribution. We assume the prior distributions $\boldsymbol{\theta}_s | \boldsymbol{\Gamma} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Gamma}_s)$ and $\mathbf{W}_s | \boldsymbol{\Sigma} \sim \text{MVN}(\mathbf{0}, \mathbf{I}_N \otimes \boldsymbol{\Sigma}_s)$ for these variables and further assume them to be statistically independent of each other, where $\text{MVN}(\boldsymbol{\mu}, \mathbf{P})$ denotes the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\mathbf{P}$, $\boldsymbol{\Gamma}_s = (\mathbf{I}_N \otimes \mathbf{S}_q) \boldsymbol{\Gamma} (\mathbf{I}_N \otimes \mathbf{S}_q^T)$, $\boldsymbol{\Sigma}_s = \mathbf{S} \boldsymbol{\Sigma} \mathbf{S}^T$ and $\boldsymbol{\Gamma} \in \mathbb{R}^{N_d N \times N_d N}$ denotes the covariance matrix of the N_d modal response at all N instants. With $\mathbf{X}_s$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ considered to be random variables, Eqn. (5) describes a linear-

normal Bayesian model, which can be conveniently represented by the probabilistic graphical model shown in Fig.1.

Fig. 1 Probabilistic graphical representation of Bayesian linear model

In this model, $Y_{rs}|X_s, \Gamma, \Sigma$ also follows the multivariate normal distribution. Furthermore, we have

$$E[Y_{rs}|X_s, \Gamma, \Sigma] = E[X_s \theta_s + W_s] = \mathbf{0}; E[\theta_s Y_{rs}^T | X_s, \Gamma, \Sigma] = E[\theta_s \theta_s^T | \Gamma] X_s^T = \Gamma_s X_s^T;$$

$$\text{Cov}[Y_{rs}|X_s, \Gamma, \Sigma] = E[Y_{rs} Y_{rs}^T | X_s, \Gamma, \Sigma]$$

$$= X_s E[\theta_s \theta_s^T | \Gamma] X_s^T + E[W_s W_s^T | \Sigma] = X_s \Gamma_s X_s^T + I_N \otimes \Sigma_s \quad (6)$$

The above equations imply that $[\theta_s; Y_{rs}]|X_s, \Gamma, \Sigma$ follows a joint multivariate normal distribution with zero-mean and covariance $[\Gamma_s \Gamma_s X_s^T; X_s \Gamma_s^T X_s \Gamma_s X_s^T + I_N \otimes \Sigma_s]$. Therefore, given Y_{rs} , the posterior distribution of $\theta_s|X_s, \Gamma, \Sigma$ is normal with mean and covariance:

$$\begin{aligned} E[\theta_s | Y_{rs}, X_s, \Gamma, \Sigma] &= \Gamma_s X_s^T [X_s \Gamma_s X_s^T + I_N \otimes \Sigma_s]^{-1} Y_{rs} \\ &= [\Gamma_s^{-1} + X_s^T (I_N \otimes \Sigma_s^{-1}) X_s] X_s^T (I_N \otimes \Sigma_s^{-1}) Y_{rs} \end{aligned} \quad (7a)$$

$$\begin{aligned} \text{Cov}[\theta_s | Y_{rs}, X_s, \Gamma, \Sigma] &= \Gamma_s - \Gamma_s X_s^T [X_s \Gamma_s X_s^T + I_N \otimes \Sigma_s]^{-1} X_s \Gamma_s^T \\ &= [\Gamma_s^{-1} + X_s^T (I_N \otimes \Sigma_s^{-1}) X_s]^{-1} = [\Gamma_s^{-1} + I_N \otimes F_s]^{-1} \end{aligned} \quad (7b)$$

In Eqn. (7b), we have introduced the Fisher information matrix $F_s = (\mathbf{S} \Phi \mathbf{S}_q^T)^T \Sigma_s^{-1} (\mathbf{S} \Phi \mathbf{S}_q^T)$ for simplicity. Eqn. (7a) provides a minimum mean-square-error estimate of θ_s , and the covariance shown in Eqn. (7b) incorporates the uncertainty inherent in the modal responses, the measurement noise and the uncertain mode shapes because of the randomness in θ_s , Σ and X_s , respectively. Note that random variables X_s , Γ and Σ depend on the structural properties and sensor quality, and that all of them remain unknown in the design stage. In order to achieve a robust sensor deployment, we need to integrate out all these uncertainties. Unfortunately, there is no closed form solution for the required high-dimension integrals, so we compute them using Monte Carlo integration, as described in Subsection 4.2.

As mentioned earlier, the primary goal of sensor placement for OMA is to identify the modal parameters, i.e. the modal frequencies, damping ratios and mode shapes. However, because the measured structural responses are nonlinearly related to the modal parameters, one cannot use a linear Bayesian model and, thereby, obtain closed-form solutions for the posterior distributions of these parameters. This makes the problem of directly optimizing sensor locations for estimating modal parameters exceedingly difficult. On the other hand, as shown in Eqn. (5), a linear relation exists between the measured structural responses and the modal responses θ_s . Furthermore, θ_s contain all the information necessary for identifying modal frequencies and damping ratios, and the mode shapes can be estimated from Eqn. (5), e.g., by the method of least squares, once θ_s are known. In fact, using θ_s as a surrogate for modal parameters in OSP has been widely accepted

since 1991[2]. Although the optimal sensor locations for estimating $\boldsymbol{\theta}_s$ are not necessarily optimal for estimating the modal parameters, the approach at least provides a suboptimal solution. This is the approach employed in this paper.

3. Methodology of OSP Based on Maximum Expected Utility

3.1 Methodology

As we have seen, in the probabilistic model for sensor placement there are uncertainties in the load, in the model parameters, and in the measurements. Consequently, the OSP is essentially a problem of making a decision under uncertainty. The Von Neumann–Morgenstern utility theorem [25] justifies the expected utility hypothesis, i.e. that the optimal decision is the one that provides maximum expected utility (MEU). Thus, MEU provides a solid mathematical criterion for solving the OSP problem.

The decision alternatives in OSP mainly include the number and location of sensors and choice of modes of interest. Let the utility function $U(d, \boldsymbol{\theta}_s)$ encode the consequences (costs) of choosing decision alternative d from a decision space $\mathcal{D}$ when the value of unknown parameters is $\boldsymbol{\theta}_s$. According to MEU, the optimal decision is given by

$$U(d^*) = \max_{d \in \mathcal{D}} \int_{\mathcal{P}} U(d, \boldsymbol{\theta}_s) p(\boldsymbol{\theta}_s) d\boldsymbol{\theta}_s \quad (8)$$

where $p(\boldsymbol{\theta}_s)$ is the posterior distribution of $\boldsymbol{\theta}_s$. The expectation with respect to the marginal distribution of $\boldsymbol{\theta}_s$ accounts for the uncertainty in $\boldsymbol{\theta}_s$; however, in the design stage, only the pre-posterior distribution of $\boldsymbol{\theta}_s | \mathbf{Y}_{rs}, \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}$ is available as listed in Section 2. Therefore, we have to consider all the possible outcomes of the measured responses $\mathbf{Y}_{rs}$, mode shapes $\mathbf{X}_s$, structure/load relevant term $\boldsymbol{\Gamma}$, and measurement noise $\boldsymbol{\Sigma}$. Eqn. (8), thus, becomes

$$\begin{aligned} U(d^*) &= \max_{d \in \mathcal{D}} \int \int \int \int_{\mathcal{P}} U(d, \boldsymbol{\theta}_s) p(\boldsymbol{\theta}_s | \mathbf{Y}_{rs}, \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) d\boldsymbol{\theta}_s p(\mathbf{Y}_{rs}, \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) d\mathbf{Y}_{rs} d\mathbf{X}_s d\boldsymbol{\Gamma} d\boldsymbol{\Sigma} \\ &= \max_{d \in \mathcal{D}} \int \int \int \int_{\mathcal{P}} U(d, \boldsymbol{\theta}_s) p(\boldsymbol{\theta}_s | \mathbf{Y}_{rs}, \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) d\boldsymbol{\theta}_s p(\mathbf{Y}_{rs} | \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) d\mathbf{Y}_{rs} p(\mathbf{X}_s) p(\boldsymbol{\Gamma}) p(\boldsymbol{\Sigma}) d\mathbf{X}_s d\boldsymbol{\Gamma} d\boldsymbol{\Sigma} \quad (9) \end{aligned}$$

where independence between $\mathbf{X}_s$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ is assumed. Hence, the marginal distribution of $\boldsymbol{\theta}_s$ is obtained by integrating out the uncertainties in $\mathbf{Y}_{rs}$, $\mathbf{X}_s$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$, thereby providing a robust sensor design against unknown structure, load and sensor properties.

3.2 Choice of Utility functions

Selecting a utility function tailored to the goals of the design problem is an important step. For the present problem of modal identification, we know that only lower modes of vibration are well excited. So, in practice, typically a few of the lower modes are selected as the targets for identification. In addition, the number of sensors sets an upper bound for the number of modes that can be identified. The number of sensors is also constrained by the cost of the sensory system. In this paper, we consider a simplified sensor placement problem: find the optimal sensor locations given the number of available sensors n and target modes $\boldsymbol{\Phi} \mathbf{S}_q$. Accordingly, three utility functions based on the quadratic loss, Shannon information, and Kullback-Leibler (K-L) divergence are investigated.

3.2.1 Quadratic Loss

Quadratic Loss corresponds to the widely used mean square error (MSE) criterion in operational modal identification. This utility function is appropriate when random phenomena are Gaussian distributed. Specifically, the expected utility function is expressed as:

$$U_1(d) = - \int \int \int \int \int_{\mathcal{P}} (\boldsymbol{\theta}_s - \bar{\boldsymbol{\theta}}_s)^T (\boldsymbol{\theta}_s - \bar{\boldsymbol{\theta}}_s) p(\boldsymbol{\theta}_s | Y_{rs}, \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) d\boldsymbol{\theta}_s p(Y_{rs} | \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) dY_{rs} p(\mathbf{X}_s) p(\boldsymbol{\Gamma}) p(\boldsymbol{\Sigma}) d\mathbf{X}_s d\boldsymbol{\Gamma} d\boldsymbol{\Sigma} \\ = - \iiint \text{tr} \{ [\boldsymbol{\Gamma}_s^{-1} + \mathbf{I}_N \otimes \mathbf{F}_s]^{-1} \} p(\mathbf{X}_s) p(\boldsymbol{\Gamma}) p(\boldsymbol{\Sigma}) d\mathbf{X}_s d\boldsymbol{\Gamma} d\boldsymbol{\Sigma} \quad (10)$$

where $\bar{\boldsymbol{\theta}}_s$ is the posterior mean of random variables $\boldsymbol{\theta}_s$. This criterion has been widely used in Bayesian experimental design, see [22, 28, 29].

3.2.2 Shannon Information

Shannon information [30] is a measure of the amount of variability associated with a random variable; it considers the probability density function instead of only the second order statistics. Since most real-life problems are governed by nonlinear equations and the random phenomenon may be quite far from being normally distributed, it may be more reasonable to use the Shannon information as the cost function in OMA. Negative Shannon information has been used by Bernardo [31] as a utility function in Bayesian experimental design and recommended by Papadimitriou [5] for OSP. For our application, this utility function has the following form:

$$U_2(d) \\ = \int \int \int \int \int_{\mathcal{P}} p(\boldsymbol{\theta}_s | Y_{rs}, \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) \ln p(\boldsymbol{\theta}_s | Y_{rs}, \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) d\boldsymbol{\theta}_s p(Y_{rs} | \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) dY_{rs} p(\mathbf{X}_s) p(\boldsymbol{\Gamma}) p(\boldsymbol{\Sigma}) d\mathbf{X}_s d\boldsymbol{\Gamma} d\boldsymbol{\Sigma} \\ \propto \iiint \ln \det[\boldsymbol{\Gamma}_s^{-1} + \mathbf{I}_N \otimes \mathbf{F}_s] p(\mathbf{X}_s) p(\boldsymbol{\Gamma}) p(\boldsymbol{\Sigma}) d\mathbf{X}_s d\boldsymbol{\Gamma} d\boldsymbol{\Sigma} \quad (11)$$

where the final result is obtained due to the joint normal distribution of $(\boldsymbol{\theta}_s, Y_{rs})$ for given values of $\mathbf{X}_s, \boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$.

There are two other ways to interpret the above selection of the utility function via the K-L divergence and mutual information. Rewrite Eqn. (11) as follows:

$$U_2(d) \propto \int \int \int \int_{\mathcal{P}} p(\boldsymbol{\theta}_s | Y_{rs}, \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) \ln \frac{p(\boldsymbol{\theta}_s | Y_{rs}, \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})}{\pi(\boldsymbol{\theta}_s)} d\boldsymbol{\theta}_s p(Y_{rs} | \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) dY_{rs} p(\mathbf{X}_s) p(\boldsymbol{\Gamma}) p(\boldsymbol{\Sigma}) d\mathbf{X}_s d\boldsymbol{\Gamma} d\boldsymbol{\Sigma} \quad (12a) \\ = \iiint \left[\int_{\mathcal{Y}} \int_{\mathcal{P}} p(\boldsymbol{\theta}_s, Y_{rs} | \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) \ln \frac{p(\boldsymbol{\theta}_s, Y_{rs} | \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})}{\pi(\boldsymbol{\theta}_s) p(Y_{rs} | \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})} d\boldsymbol{\theta}_s dY_{rs} \right] p(\mathbf{X}_s) p(\boldsymbol{\Gamma}) p(\boldsymbol{\Sigma}) d\mathbf{X}_s d\boldsymbol{\Gamma} d\boldsymbol{\Sigma} \quad (12b)$$

Eqn. (12a) is the expected K-L divergence [30] between the posterior and prior distributions of parameters $\boldsymbol{\theta}_s$. $U_2(d)$ and the K-L divergence achieve their maxima under the same decision d because the prior distribution $\pi(\boldsymbol{\theta}_s)$ does not depend on d . K-L divergence is a measure of distance between two probability distributions, and here, maximizing it implies extracting more information from the data other than the prior. Eqn. (12b) expresses the expected mutual information [30] of parameters $\boldsymbol{\theta}_s$ and data Y_{rs} conditioned on $\mathbf{X}_s, \boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$. Intuitively, mutual information measures the information that $\boldsymbol{\theta}_s$ and Y_{rs} share; it measures the amount of uncertainty reduction in one set of variable when the other set is known. Hence, maximizing the expected

gain in the Shannon information is equivalent to maximizing the dependence of $\boldsymbol{\theta}_s$ on data $\mathbf{Y}_{rs}$. Besides these interpretations, Eqn. (11) can also be derived from other utility functions, see [32, 33].

3.2.3 K-L Divergence

As described above, the K-L divergence is a measure of distance between two probability distributions. Li [8] has introduced the K-L divergence criterion for OSP. Unlike the preceding utility functions that were focused on OMA, the proposition here is more test-oriented: the perfect information on modal parameters $\boldsymbol{\theta}_s$ can be extracted by placing sensors everywhere in the structure; so, the optimal configuration for a given set of $n < N_d$ sensors should provide information that is as close as possible to that of full sensor set. Contrary to the value of information [34] that quantifies the reduction in expected loss (negative utility) due to the availability of the information, K-L divergence measures the reduction in utility by removal of sensors from the full configuration. Since for the N_d sensor case $\mathbf{S} = \mathbf{I}_{N_d}$ and $\mathbf{S}_q$ doesn't change, hereafter the subscript s on all symbols for this case are removed except for $\boldsymbol{\theta}_s$ and $\boldsymbol{\Gamma}_s$.

Let $p(\boldsymbol{\theta}_s | \mathbf{Y}_{rs}, \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$ denote the posterior distribution of $\boldsymbol{\theta}_s$ from n sensors and $p(\boldsymbol{\theta}_s | \mathbf{Y}_r, \mathbf{X}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$ denote the posterior distribution of $\boldsymbol{\theta}_s$ from N_d sensors. Selecting the negative K-L divergence between $p(\boldsymbol{\theta}_s | \mathbf{Y}_{rs}, \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$ and $p(\boldsymbol{\theta}_s | \mathbf{Y}_r, \mathbf{X}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$ as the utility function, we have

$$\begin{aligned} U_3(d) &= - \iiint \int \int_{\mathcal{P}} p(\boldsymbol{\theta}_s | \mathbf{Y}_{rs}, \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) \ln \frac{p(\boldsymbol{\theta}_s | \mathbf{Y}_{rs}, \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})}{p(\boldsymbol{\theta}_s | \mathbf{Y}_r, \mathbf{X}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})} d\boldsymbol{\theta}_s p(\mathbf{Y}_r | \mathbf{X}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) d\mathbf{Y}_r p(\mathbf{X}) p(\boldsymbol{\Gamma}) p(\boldsymbol{\Sigma}) d\mathbf{X} d\boldsymbol{\Gamma} d\boldsymbol{\Sigma} \\ &\propto - \iiint \left\{ \frac{1}{2} \ln \det[\boldsymbol{\Gamma}_s^{-1} + \mathbf{I}_N \otimes \mathbf{F}_s] + \text{tr}\{[\boldsymbol{\Gamma}_s^{-1} + \mathbf{I}_N \otimes \mathbf{F}_s]^{-1} [\boldsymbol{\Gamma}_s^{-1} + \mathbf{I}_N \otimes \mathbf{F}]\} \right\} p(\boldsymbol{\Gamma}) p(\mathbf{X}) p(\boldsymbol{\Sigma}) d\mathbf{X} d\boldsymbol{\Gamma} d\boldsymbol{\Sigma} \quad (13) \end{aligned}$$

The derivation of this utility function is not as straightforward as the previous two; it is described in Appendix A. This utility function is related to the preceding two utility functions: the first term inside the large curly bracket is one half of $U_2(d)$ and the second term can be regarded as a normalized version of $U_1(d)$. As can be seen in the examples below, the second term dominates the utility function $U_3(d)$.

Besides the motivation behind the K-L divergence, there is another meaningful interpretation. Recall $\mathbf{Y}_r = [\mathbf{Y}_{rs}^T, \mathbf{Y}_{rn}^T]^T$, where $\mathbf{Y}_{rn}$ denotes the response at the non-selected DOFs. Eqn. (13) can be equivalently written as

$$\begin{aligned} U_3(d) &= - \iiint \int \left[\int p(\boldsymbol{\theta}_s | \mathbf{Y}_{rs}, \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) p(\mathbf{Y}_{rn} | \mathbf{Y}_{rs}, \mathbf{X}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) \ln \frac{p(\boldsymbol{\theta}_s | \mathbf{Y}_{rs}, \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) p(\mathbf{Y}_{rn} | \mathbf{Y}_{rs}, \mathbf{X}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})}{p(\boldsymbol{\theta}_s, \mathbf{Y}_{rn} | \mathbf{Y}_{rs}, \mathbf{X}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})} d\boldsymbol{\theta}_s d\mathbf{Y}_{rn} \right] \\ &\quad \cdot p(\mathbf{Y}_{rs} | \mathbf{X}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) d\mathbf{Y}_{rs} p(\mathbf{X}) p(\boldsymbol{\Gamma}) p(\boldsymbol{\Sigma}) d\mathbf{X} d\boldsymbol{\Gamma} d\boldsymbol{\Sigma} \quad (14) \end{aligned}$$

in which the term inside the square bracket is the K-L divergence of the product of conditional distributions $p(\boldsymbol{\theta}_s | \mathbf{Y}_{rs}, \mathbf{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$ and $p(\mathbf{Y}_{rn} | \mathbf{Y}_{rs}, \mathbf{X}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$ and the joint distribution $p(\boldsymbol{\theta}_s, \mathbf{Y}_{rn} | \mathbf{Y}_{rs}, \mathbf{X}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$. Maximizing utility $U_3(d)$ means minimizing the above K-L divergence, i.e. making the joint distribution of $\boldsymbol{\theta}_s$ and $\mathbf{Y}_{rn}$ given $\mathbf{Y}_{rs}$ approximately equal to their product. Indeed, this assures conditional independence of $\boldsymbol{\theta}_s$ and $\mathbf{Y}_{rn}$, given $\mathbf{Y}_{rs}$. Therefore, this criterion makes the modal parameters and the unmeasured responses as independent as possible, given the measured responses.

3.3 Elicitation of Prior Information

Eliciting prior distributions is an important step for robust optimal sensor placement, because the terminal decision may strongly depend on the prior information [35]. Here, we need to assign prior distributions to $\mathbf{X}_s$, $\mathbf{\Gamma}$ and $\mathbf{\Sigma}$. The source of prior information can be the nominal structural model, earlier related experiments, observational studies, or subjective belief. Nominal model-based prior elicitation, e.g. based on a finite element model of the structure, is preferred in civil engineering.

The prior distribution of $\mathbf{X}_s$ represents the uncertainty in our characterization of the target structure. A finite element model is suitable to provide this information, in which the nominal value from the design drawings serves as the mean and the covariance matrix of input parameters is specified by expert knowledge to represent the confidence in the nominal model. Since $\mathbf{X}_s$ only contains the mode shapes, we can alternatively randomize the mode shape matrix, which perhaps is more efficient because repeated eigenvalue analysis is avoided. In this paper, we choose to use a finite element model due to its intuitiveness, although more computation is involved.

Random variables $\mathbf{\Gamma}$ stand for the covariance matrix of the modal responses. The uncertainty in $\mathbf{\Gamma}$ comes from two sources: the structural model and the load pattern. Our focus in this paper is on sensor placement for operational modal analysis (OMA). A basic assumption in OMA is that the input is band-limited white noise (BWN). Consequently, only the magnitude and bandwidth of the power spectrum density (PSD) of the BWN need to be specified. In fact, we only need to specify this magnitude relative to the magnitude of the measurement noise. This can be interpreted as the signal-to-noise ratio, which is an important factor for designers to consider. As for the bandwidth, the usual assumption is that it includes at least all the modes of interest. Measurements on site can be used to guide the specification of these parameters in practice. We address detailed specification of $\mathbf{\Gamma}$ for the BWN excitation in the next section.

$\mathbf{\Sigma}$ characterizes the uncertainties in the measurement system, which may include sensor error, transmission error, transducer error and computer error. Strictly speaking, these errors can be spatially correlated, but for the ease of mathematical treatment, they are usually treated as independent and identically distributed random variables. Here, we will employ this approximation to assume $\mathbf{\Sigma} = \sigma^2 \mathbf{I}_n$.

4. Computational Issues

4.1 Modal Responses under BWN Excitation

Measurements of structural responses, such as displacement, velocity and acceleration, are used for modal identification. For computational efficiency, closed form solutions of the covariance matrix of these responses to a BWN excitation are desired. Crandall and Mark [36] derived the variance of the displacement response under BWN excitation. However, to our knowledge, closed form solutions of the variances and covariances of the velocity and acceleration responses under BWN excitation are not available. Here, we derive these results by use of partial fraction integration. A similar approach was used by Der Kiureghian [37] for a stationary excitation and Harichandran [38] for a propagating excitation.

Assume each pair of the elements of the forces $\mathbf{f}$ or $\ddot{\mathbf{u}}_g$ in the modal equation of motion in Eqn. (2) are totally incoherent and each is modeled as a BWN with a PSD magnitude of $\Phi_{0,k}$ and bandwidth of $\omega_{c,k}$. The cross-PSD matrix of the input force then has the elements

$$\Phi_{kl} = \begin{cases} \Phi_{0,k}\delta_{kl}, & \text{if } -\omega_{c,k} \leq \omega \leq \omega_{c,k} \\ 0, & \text{elsewhere} \end{cases}; k, l = 1, 2, \dots, N_f \quad (15)$$

in which δ_{kl} is the Kronecker delta. Under this input PSD, the covariances of the i th and the j th modal responses can be calculated by

$$\Gamma_{q_i^{(e)} q_j^{(f)}}(\tau) = (-1)^f I^{e+f} \sum_{k=1}^{N_f} \gamma_{ik} \gamma_{jk} \Phi_{0,k} \sum_{l=1}^4 \text{Res}(\Omega_l) \int_{-\omega_{c,k}}^{\omega_{c,k}} \frac{\omega^{e+f} e^{I\omega\tau}}{\omega - \Omega_l} d\omega, e, f = \{0, 1, 2, 3, 4\} \quad (16)$$

where $q_i^{(e)}$ represents the e th derivative of the i th modal displacement with respect to time, γ_{ik} stands for the i th modal participation factor with respect to the k th BWN excitation, $I = \sqrt{-1}$ is the imaginary unit, Ω_l and $\text{Res}(\Omega_l)$ are the l th pole and corresponding residual of $Q(\omega) = H_i(\omega)H_j^*(\omega)$, in which $H_i(\omega)$ is the frequency-response function (FRF) of mode i , and an asterisk denotes the complex conjugate. The detailed derivation and the closed form of the integral in Eqn.(16) are given in Appendix B.

4.2 Calculation of Utility Function

Recall that we need to evaluate high dimensional integrals over $\mathbf{X}_s, \mathbf{\Gamma}$ and $\mathbf{\Sigma}$ for calculation of the utility functions $U_1(d)$, $U_2(d)$ and $U_3(d)$. Obtaining closed-form solutions of these integrals or using numerical integration are impractical. Hence, we select to use the Monte Carlo sampling (MCS) technique to obtain an approximate solution. MCS is relatively simple for our problem because independent priors are assumed for random variables representing the excitation, the structural parameters (masses and stiffnesses) and the measurement error. For each realization of structural parameters, the mode shapes $\mathbf{X}_s$ are computed.

Let N_s denote the sample size. The utility functions are computed as follows:

$$U_1(d) \cong -\frac{1}{N_s} \sum_{i=1}^{N_s} \text{tr} \{ [\mathbf{\Gamma}_{s,i}^{-1} + \mathbf{I}_N \otimes \mathbf{F}_{s,i}]^{-1} \} \quad (17a)$$

$$U_2(d) \cong \frac{1}{N_s} \sum_{i=1}^{N_s} \ln \det [\mathbf{\Gamma}_{s,i}^{-1} + \mathbf{I}_N \otimes \mathbf{F}_{s,i}] \quad (17b)$$

$$U_3(d) \cong -\frac{1}{N_s} \sum_{i=1}^{N_s} \left\{ \frac{1}{2} \ln \det [\mathbf{\Gamma}_{s,i}^{-1} + \mathbf{I}_N \otimes \mathbf{F}_{s,i}] + \text{tr} [\mathbf{\Gamma}_{s,i}^{-1} + \mathbf{I}_N \otimes \mathbf{F}_{s,i}]^{-1} [\mathbf{\Gamma}_{s,i}^{-1} + \mathbf{I}_N \otimes \mathbf{F}_i] \right\} \quad (17c)$$

where the subscript i on each term indicates the i th simulated value. Another concern arises from the high dimension of $\mathbf{\Gamma}_s$, which equals the product of the sampling frequency f_s , time duration T_d , and the number of target modes m . If this dimension is too high, the inverse and determinant operations on $\mathbf{\Gamma}_s$ will be unfeasible. If $\mathbf{\Gamma}$ is n by n , the memory required to store $\mathbf{\Gamma}_s$ is $O(n^2)$ and the number of floating-point operations (Flops) required for matrix inverse and determinant calculations is about $O(n^3)$. Therefore, for large n , it is necessary to develop an approximate solution method. Considering the block Toeplitz structure of the covariance matrix $\mathbf{\Gamma}_s$, there are two possible strategies available in the literature: the nearest block circulant matrix approximation [39] and the nearest Kronecker product approximation [40]. These approximations reduce the number of operations from $O(n^3)$ to $O(n^2 \log n)$ and $O(n_1^3 + n_2^3)$, respectively, where $n_1 + n_2 = n$. Because of the possibility of high reduction in the number of operations and its appealing algebra [41], we choose the nearest Kronecker product approximation. Analysis of the error in this approximation can be found in [42]. For the sake of completeness, a detailed treatment of this method is reported in Appendix C, where it is shown that we only need to evaluate a block diagonal matrix multiplication and get its trace in $U_1(d)$ and $U_3(d)$; as for $U_2(d)$, only a summation over diagonal elements of a diagonal matrix is needed. Among the three, evaluation of $U_2(d)$ is most efficient.

5. Application

5.1 Shear-type Building

The first example investigated is a 16-DOF shear-type building simplified as a mass-spring system, shown in Fig. 2. The floor masses $m_1 \sim m_{16}$ are modeled as joint lognormally distributed random variables with different logarithmic means λ_m but identical coefficient of variation (C.O.V) δ_m and cross-correlation coefficients $\rho_{\ln m}$. The story stiffnesses are modeled in a similar way. Modal damping ratios are assumed to be identically and independently lognormally distributed with a larger C.O.V. The building is subjected to base motion and wind load, both of which are modeled as independent BWN processes with the same PSD magnitude of Φ_0 and bandwidth, $\omega_c/2\pi = 5 \text{ Hz}$. Elements of the load coefficient vector of the wind load are assumed to be independent normal random variables with mean $P = 10^6 \times 0.616 \left(\frac{z}{10}\right)^{0.44}$ [43] and 5% C.O.V, in which $z = kh$, $k = 0, \dots, 16$, and $h = 3 \text{ m}$ is the floor height. The same input to measurement noise ratio (INR) $2\Phi_0\omega_c/\sigma^2$ is applied for the measurement of acceleration, velocity and displacement, and is modeled by a uniform distribution bounded between 1 and 100. This quantity is dimensionless for measurement of acceleration, but has units of sec^{-2} and sec^{-4} for velocity and displacement, respectively. Although acceleration is usually measured more accurately than displacement and velocity, for the sake of simplicity in this example values of INR for all three are assumed to be in the same level. All these variables are defined in Tab.1. We also set the sampling frequency as $f_s = 50 \text{ Hz}$ and duration as $T_d = 60 \text{ s}$ and select the MCS sampling size as $N_s = 1000$. Through a number of tests, we determined that a sample of size of $N_s = 1000$ is sufficient to ensure convergence of the estimates based on MCS within a coefficient of variation smaller than 0.01.

The lowest four modes of the structure are of interest. The histograms of the natural frequencies for the 1000 realizations are shown in the upper part of Fig. 3. The mean mode shapes (black lines) and 2 standard deviation intervals are shown in the bottom part of Fig. 3. The distributions of the natural frequencies appear to be lognormal. This is because the floor masses and story stiffnesses were assigned joint lognormal distributions. The fundamental frequency has a sharper distribution than the other frequencies. Similarly, the variability in the mode shapes increases with the order number. The uncertainties in mode shapes influence the OSP through $\mathbf{X}_s$ shown in Eqn.(5), but the effects of modal frequencies on OSP are not straightforward. Modal frequencies influence the unknown modal response $\boldsymbol{\theta}_s$ and, therefore, the mean and covariance matrix of $\boldsymbol{\theta}_s$. Also, recall from Eqn.(16) that computation of the prior covariance matrix $\boldsymbol{\Gamma}_s$ depends on modal frequencies. Therefore, the influence of uncertainties in modal frequencies on the OSP is through $\boldsymbol{\Gamma}_s$.

The auto-covariance function (ACF) of the 1st modal acceleration and the cross-covariance function of the 1st and 2nd modal accelerations for one MCS realization are illustrated in Fig. 4. Results obtained from the closed-form solution and from numerical integration match, verifying the closed-form solution. Another distinct observation is the large magnitude of the variance relative to the Fisher information $\mathbf{F}_s$ which is less than of order 10^2 in the example, implying large uncertainty in the modal acceleration responses. Given the large variance, it is expected that the inverse of the covariance matrix $\boldsymbol{\Gamma}_s$ will not make a significant contribution to the selected utility functions when acceleration measurements are made.

Figure 2. 16 DOFs shear-type building

Tab.1 Probabilistic model for example structure

Variable	Distribution	Mean	C.O.V	Correlation
$m_1 \sim m_4$	Joint Lognormal	5×10^5	0.05	$\rho_{\ln m} = 0.6$
$m_5 \sim m_8$		4×10^5		
$m_9 \sim m_{12}$		3×10^5		
$m_{13} \sim m_{16}$		2×10^5		
$k_1 \sim k_4$	Joint Lognormal	1.6×10^8	0.1	$\rho_{\ln k} = 0.6$
$k_5 \sim k_8$		1.5×10^8		
$k_9 \sim k_{12}$		1.4×10^8		
$k_{13} \sim k_{16}$		1.3×10^8		
$\xi_1 \sim \xi_{16}$ (%)	Joint Lognormal	2	0.3	Independent
$2\Phi_0\omega_c/\sigma^2 \omega_c$	Uniform(1,100)	50.5	0.515	Independent

Fig. 3 Variation of Modal Parameters

Fig. 4 ACF and CCF for one MCS realization

5.1.1 Non-Bayesian Placement of Sensors

In a non-Bayesian approach, i.e., without prior information, $\mathbf{F}_s^{-1} = \mathbf{0}$ and there is no difference in deploying displacement meter, velocimeter or accelerometer for modal identification. As a consequence, different types of sensors will occupy the same locations in a design scenario of mixed sensor placement. Therefore, for comparison, we first deploy sensors based on the nominal model and the probabilistic model considering the uncertainty in mode shapes only. Using the backward sequential method [5], the optimal configuration of sensors for the three utility functions is shown in Figs.5 and 6. Because the Fisher information matrix becomes singular if the number of sensors is less than the number of modes, the placement starts from 4 sensors. The distinction between Figs.5 and 6 lies in the consideration of the uncertainty in mode shapes in the latter. Comparing the two sets of figures, we see that the two methods give almost identical configurations for small number of sensors, but discrepancy between the results occurs for moderate number of sensors. The similarity within these configurations can be partly explained by the unbiasedness assumption on the nominal model, but it is hard to find a more reasonable model assumption in the design stage. The quadratic loss utility function $U_1(d)$ provides the same configuration for both scenarios, and this can be attributed to the linearity of the trace and mean operations; while the utility function $U_3(d)$ based on the K-L divergence is the most sensitive to the uncertainty in mode shapes, because it includes the combined nonlinearity of determinant and normalized trace terms.

Fig. 5 Configuration of sensors based on nominal model
Left to right: $U_1(d), U_2(d), U_3(d)$.

Fig. 6 Configuration of sensors based on probabilistic model considering the uncertainty in mode shapes
Left to Right: $U_1(d)$, $U_2(d)$, $U_3(d)$.

5.1.2 Placement of Displacement Meters

Next, we apply the proposed Bayesian method for displacement meters with the assumed prior distributions. The results are shown in Fig. 7. Note that with this method we can deploy fewer sensors than the number of modes. This is because the covariance matrix, here consisting of not only the Fisher information matrix but also the prior covariance matrix, is not singular. Comparing with Fig. 6, we see that the optimal configurations of sensors are influenced by the prior information. For example, more sensors are placed near the top of the structure and DOF 4 is now more preferred than DOF 3. The optimal sensor configurations are quite different under the three utility functions: the Shannon information is more influenced by prior information and tends to cluster sensors on the top, while the K-L divergence and quadratic loss give more similar configurations and distribute sensors more evenly. This is because both the latter utility functions involve similar trace terms, to which the FIM contributes more in this example.

Fig. 7 Configuration of displacement meters: a) $U_1(d)$; b) $U_2(d)$; c) $U_3(d)$.

In order to check how the INR influences the optimal configuration of sensors, the configurations of displacement meters based on $U_2(d)$ are shown in Fig. 8 under different values of INR. When the INR is extremely low (INR = 1), $\mathbf{\Gamma}_s^{-1}$ contributes more to the determinant, and all sensors cluster in the top of the building where the kinetic energy is higher. When the INR is high, the influence of $\mathbf{\Gamma}_s^{-1}$ diminishes and the FIM becomes dominant so that distributions of sensors are more even. Similar results are observed for the other two utility functions, but not so dramatically. Thus, the INR setting connects the kinetic energy-based and the FIM-based approaches, and a balanced configuration can be achieved by setting a reasonable INR.

Fig. 8 Relation between configurations of sensors and INR.

5.1.3 Placement of Velocimeters

Fig. 9 shows the placement of velocimeters incorporating the prior covariance of modal velocity responses. Comparing with Figs. 6 and 9, we see that the prior information changes the configuration of velocime-

ters, but less severely than for the case of displacement meters. The reason is that the modal velocity has larger uncertainty than that of modal displacement, so that the inverse of the prior covariance matrix $\mathbf{\Gamma}_s^{-1}$ contributes less in determining the optimal locations. Optimal configurations of sensors differ for different utility functions, but DOFs 3, 6 or 7, 11 and 16 are selected in all configurations. Again, the K-L divergence yields a more even distribution of velocimeters.

Fig. 9 Configuration of velocimeters: a) $U_1(d)$; b) $U_2(d)$; c) $U_3(d)$

5.1.4 Placement of Accelerometers

Fig.10 shows the placement of accelerometers, while incorporating the prior information. The results for 4 or more sensors are almost identical to those in Fig. 6. This is because the prior information does not contribute much to the utility function, as explained in the preceding section. Mathematically, this is because the inverse of $\mathbf{\Gamma}_s$ is negligible compared to the other terms in the utility functions. This observation justifies the use of non-Bayesian methods in the placement of accelerometers when the INR is high. Different utility functions give distinct configurations, but in all cases clusters occur around DOFs 3, 6 or 7, 11 and 16.

5.1.5 Mixed Placement of Displacement Meters and Accelerometers

Next, we consider a mixture of displacement meters and accelerometers for modal identification. The non-Bayesian method does not distinguish between different types of sensors, hence it cannot solve this problem. In this paper, we consider a simple example of placing one displacement meter and three accelerometers to identify four modes. More general cases, e.g. optimizing the type and position of a fixed number of sensors, can be solved in our framework; however, computational difficulties may be encountered in the optimization scheme and more work is needed before such a general problem can be solved. For our simple example, solutions obtained by cross entropy optimization [8] are listed in Tab. 2 for the three utility functions. The utility function based on Shannon information results in the displacement sensor and one accelerometer occupying the same location. This could be due to our neglecting the correlations between measurement errors, which has a larger influence in the case of the utility function based on Shannon information. The collocated displacement sensor and accelerometer may provide redundant information in identifying mode shapes, but this collocation may be desirable in practice in order to accurately identify the frequencies and damping ratios of lower modes from the low-frequency displacement data and the frequencies and damping ratios of higher modes from the high-frequency acceleration data, even though our algorithm cannot explicitly account for these aspects. The other two utility functions give identical configurations: the displacement sensor is placed at the top of the building, where the signal-to-noise ratio is highest, and the accelerometers are placed evenly to improve the identifiability of the mode shapes.

Fig. 10 Configuration of accelerometers: a) $U_1(d)$; b) $U_2(d)$; c) $U_3(d)$
Tab.2 Sensor location in mixed scenario

Utility functions	Locations	
	Displacement meter	Accelerometer
$U_1(d)$	16	[5, 11, 16]
$U_2(d)$	16	[4, 9, 13]
$U_3(d)$	16	[4, 9, 13]

5.2 Guangzhou TV Tower

The Guangzhou TV Tower (Fig. 11a), located in Guangzhou, China, is a super tall structure consisting of a 454 meters main tower and a 156 meters antenna mast. It has been developed as a benchmark problem for structural health monitoring [27], with optimal sensor placement being Task III in the program. Many researchers [44,45] have used this structure as a test bed for investigating sensor placement algorithms. Here, we use this structure to illustrate the applicability of the proposed method.

A reduced finite element model (FEM) of the structure (Fig. 11b) can be downloaded from <http://www.cse.polyu.edu.hk/benchmark/>. This FEM includes 27 elements for the main tower and 10 elements for the antenna mast. Nodes are numbered from 1 at the fixed base to 38 at the free top end. In order to obtain a robust design, the structural and force parameters are first randomized. Since the nominal element stiffness matrices provided are not positive semidefinite, the way used in Example 1 to generate the structural stiffness cannot guarantee its positive definiteness. In this paper, we apply a two-step generating method: firstly, generate stiffness matrices modeled by multivariate normal distribution with means presented by the nominal values and C.O.V. 0.05; secondly, calculate the mean and covariance of the generated stiffness matrices and use these parameters to generate stiffness matrices following the Wishart distribution [46]. Element masses are modeled by independent lognormal distributions with means represented by the nominal values and a C.O.V. of 0.02. Classical damping is assumed, and modal damping ratios are assigned the same distribution as in the first example. The input force model is identical with that in the first example, and both the ground motion and wind load are considered. The other parameters, i.e., the sampling frequency, the time duration, and the number of Monte Carlo samples are identical to those assumed for first example.

The Guangzhou TV Tower is a typical primary-secondary structure. The main tower is made of a reinforced concrete interior tube and steel external tube, while the antenna mast is made of steel. Relative to the main tower, the antenna mast is lighter and more flexible. As a result, the mode shapes of the combined system are expected to have much larger amplitudes for DOFs associated with the antenna mast than those associated with the main tower. Furthermore, it is expected that the antenna mast will have relatively insignificant influence on the dynamic response of the tower. In the following, we consider two cases for sensor placement: one is to retain the antenna DOFs as possible locations for sensors, in which case we obtain the OSP solution to identify the modes of the combined system, and the other is to disallow placement of sensors at the DOFs of the antenna, in which case the OSP solution aims at identifying the dominant modes of the tower. Because rotational DOFs are difficult to measure, only translational DOFs in X and Y directions are considered for possible sensor installation. Furthermore, only the placement of accelerometers is considered since that is the most commonly used sensor for modal identification.

5.2.1 Main Tower with Antenna Mast

In this case, the first 15 modes are selected as the modes of interest. The sensor configurations under different utility functions are shown in Fig. 12. It is seen that more than half the sensors are clustered in the antenna mast, showing its importance in modeling the dynamical behavior of the combined system. Because for each mode shape the amplitudes at DOFs associated with the antenna mast are much larger than those of the main tower, these locations give a relatively large FIM. By checking the inverse of the covariance matrix of the modal accelerations, it is found that the magnitude of this matrix is much smaller than that of the FIM, so that the FIM dominates the optimization. In addition, due to the asymmetry of the structure, the configurations of accelerometers are not identical in the X and Y directions; indeed, even the number of accelerometers in many cases are not the same. There are significant differences in the configurations of sensors under different utility functions, but it is hard to say which one is better unless more evaluation criteria are included.

Fig. 11 Guangzhou TV Tower: a) Overview; b) Reduced FEM

Fig. 12 Configuration of accelerometers: a) $U_1(d)$; b) $U_2(d)$; c) $U_3(d)$
 ○: X direction; □: Y direction

5.2.2 Main Tower without Antenna Mast

If the engineer is more concerned about the modal properties of the main tower, which are more complex and uncertain than those of the antenna mast, then it is advisable to exclude the antenna mast from candidate sensor locations. Therefore, we now consider placing sensors only on the main tower.

There are 11 vibration modes of the main tower in the first 15 modes of the combined system, and they are selected as the target modes. It is noted that a FEM of only the main tower was used for OSP in [44], but this way of modeling is not appropriate because the mode shapes of the main tower are influenced by the antenna mast. In fact, a different structure is modeled if the antenna mast is totally ignored. Hence, here, the FEM of the full structure is used, but only the DOFs corresponding to the main tower are selected as the potential locations of sensors. The optimal configurations of sensors are listed in Fig. 13 for 6, 12 and 18 accelerometers. Some similarities exist between the results for this case and the case considering the antenna mast for the tower DOFs. In particular, the number and locations of accelerometers are not identical in the X and Y directions because of the asymmetry of the structure. However, some configurations appear to be more regular in this case, e.g., the placement of 12 accelerometers based on the K-L Divergence utility function. An interesting observation based upon all examples in the paper is that the sparsity of the configuration depends on the relation between the number of available sensors n and the number of modes of interest m : the configuration does not cluster together if n is not much larger than m ; otherwise clusters may occur when n increases.

The relation between values of the utility functions and the number of sensors is shown in Fig. 14. It can be seen that the utility functions monotonically increase with the number of sensors. The rate of increase is particularly high when the number of sensors is smaller than the number of modes of interest. This implies that the number of sensors to be placed should be at least equal to or greater than the number of modes to be identified. When the antenna mast is removed as a potential sensor location, significant drops occur in all utility function values. Obviously, optimal configurations without the antenna mast are inferior to those considering all the DOFs. When the number of sensors is small, the values of utility function $U_1(d)$ with or

without the antenna mast are only slightly different. This is because $\mathbf{F}_s^{-1}$ dominates the trace operation for this utility function, not the FIM term where the difference between the two scenarios lies. Utility function $U_3(d)$ shows the largest drop in value when removing the antenna mast as a potential sensor location. This is because all available DOFs are used as the full set configuration in the process of computing $U_3(d)$.

Fig. 13 Configuration of accelerometers without antenna mast: a) $U_1(d)$; b) $U_2(d)$; c) $U_3(d)$
 $\circ$: X direction; $\square$: Y direction

Fig. 14 Relations between utilities and numbers of sensors: a) $U_1(d)$; b) $U_2(d)$; c) $U_3(d)$

6. Conclusions

A Bayesian probabilistic model for sensor placement considering model and load uncertainty and measurement error is formulated. The maximum expected utility theory is employed to select optimal sensor locations for modal response estimation, and this optimal configuration works as a suboptimum for the OSP in OMA. Three utility functions are considered, each based on a different motivation. The prior covariance of modal responses is theoretically derived, and its nearest Kronecker product approximation is developed for fast evaluation of the utility functions. Conclusions derived from two cases studies are as follows:

- (1) The OSP is essentially a decision problem under uncertainty. The proposed framework consisting of Bayesian statistics and the maximum expected utility criterion provides a powerful tool to address all relevant uncertainties in the sensor placement problem and to incorporate all available prior information. By specifying a suitable utility function, a robust optimal configuration of sensors can be obtained.
- (2) When accounting for prior information, the optimal placement configuration of displacement meter, velocimeter and accelerometer can be different; indeed, mixed sensor placement becomes possible. Since the prior covariance matrix of modal acceleration has a large magnitude, the placement of accelerometer is less influenced by the prior information. This justifies the commonly used mode-shape-based accelerometer placement, since the prior information of modal acceleration does not strongly influence the final result.
- (3) The magnitude of INR has a great influence on the optimal configuration of sensors. If it is small, the magnitude of prior covariance of modal responses is also small; the result is that sensors are placed at positions with high kinetic energy; if the magnitude of the INR is large, the prior information could be ignored, and the configuration is fully determined by the FIM. Thus, the magnitude of the INR determines the balance between the kinetic energy and Fisher information.
- (4) Model uncertainties considered in the example applications do not significantly influence the optimal configuration of sensors, especially when the number of sensors available is small. This observation is partly due to the fact that the nominal structural model is assumed to be unbiased; the case of a biased model is left for future study.
- (5) Three utility functions are introduced based on quadratic loss, Shannon information and K-L divergence. As expected, different utility functions yield different configurations of sensors. The utility function based on Shannon information is most influenced by the prior information, and its computation is the most efficient;

the K-L divergence based utility function gives relatively even distribution of sensors. As for how to select the utility function, more evaluating criteria are needed.

(6) The number of sensors placed should be at least equal to or more than the number of modes of interest. In this paper we assume both numbers are predetermined. In fact, these two parameters can be also incorporated into the maximum expected utility framework by specifying suitable utility functions. This will be addressed in a future work.

7. Acknowledgements

Support from the National Science Foundation under Grant No. CMMI-1130061 is gratefully acknowledged.

References

- [1] K. Yuen, Bayesian Methods for Structural Dynamics and Civil Engineering, John Wiley & Sons (Asia) Pte Ltd, Singapore, 2010.
- [2] D.C. Kammer, Sensor placement for on-orbit modal identification and correlation of large space structures, *Journal of Guidance, Control and Dynamics*, 14(1991) 251-259.
- [3] F.E. Udwadia, Methodology for optimum sensor locations for parameter identification in dynamic systems, *Journal of Engineering Mechanics*, 120(2)(1994) 368-390.
- [4] S. Borguet, O. Léonard, The Fisher information matrix as a relevant tool for sensor selection in engine health monitoring, *International Journal of Rotating Machinery*, 2008(2008) 1-10.
- [5] C. Papadimitriou, Optimal sensor placement methodology for parametric identification of structural systems, *Journal of Sound and Vibration*, 278(2004) 923-947.
- [6] C. Papadimitriou, G. Lombaert, The effect of prediction error correlation on optimal sensor placement in structural dynamics, *Mechanical Systems and Signal Processing*, 28(2012) 105-127.
- [7] I. Trendafilova, W. Heylen, H. Van Brussel, Measurement point selection in damage detection using the mutual information concept, *Smart Materials & Structures*, 10(3)(2001) 528-533.
- [8] B. Li, J. Ou, Optimal sensor placement for structural health monitoring based on K-L divergence, in: *Proceeding of Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures*, New York, USA, 2013.
- [9] B. Li, Information theoretic optimal sensor placement in structural health monitoring, Master Thesis, Dalian University of Technology, Dalian, China, 2012.
- [10] M. Papadopoulos, E. Garcia, Sensor placement methodologies for dynamic testing, *AIAA Journal*, 36(1998) 256-263.
- [11] W. Heylen, S. Lammens, P. Sas, *Modal Analysis Theory and Testing*, Katholieke Universiteit Leuven, Leuven, Belgium, 1998.
- [12] N. Debnath, A. Dutta, S.K. Deb, Placement of sensors in operational modal analysis for truss bridges, *Mechanical Systems and Signal Processing*, 31 (2012) 196-216.
- [13] J. Penny, M. Friswell, S. Garvey, Automatic choice of measurement locations for modal survey test, *AIAA Journal*, 32(1994) 407-414.
- [14] D.J. Ewins, *Modal Testing: Theory, Practice and Application*, Research Studies Press, Baldock, UK, 2000.
- [15] W. Waldruff, D. Dochain, S. Bourrel, A. Magnus, On the use of observability measures for sensor location in tubular reactor, *Journal of Process Control*, 8(1998) 497-505.
- [16] J.A. Wilson, S.Y. Guhe, Observability matrix condition number in design of measurement strategies, *Computer Aided Chemical Engineering*, 20(2005) 397-402.
- [17] D. Li, Y. Zhang, L. Ren, H. Li, Sensor deployment for structural health monitoring and their evaluation, *Advances in Mechanics*, 41(1)(2011) 39-50.
- [18] D. Li, H. Li and C.P. Fritzen, Load dependent sensor placement method: theory and experimental validation, *Mechanical Systems and Signal Processing*, 31(2012) 217-227.

- [19] M. Brehm, V. Zabel, C. Bucher, Optimal reference sensor positions using output-only vibration test data, *Mechanical Systems and Signal Processing*, 41(2013) 196-225.
- [20] P. Vinot, C. Cogan, V. Cipolla, A robust model-based test planning procedure, *Journal of Sound and Vibration*, 288(2005) 571–585.
- [21] R. Castro-Triguero, S. Murugan, R. Gallego, M.I. Friswell, Robustness of optimal sensor placement under parametric uncertainty, *Mechanical Systems and Signal Processing*, 41(2013) 268–287.
- [22] E. Heredia-Zavoni, L. Esteva, Optimal instrumentation of uncertain structural systems subject to earthquake ground motions, *Earthquake Engineering & Structural Dynamics*, 27(4)(1998) 343-362.
- [23] K.V. Yuen, L.S. Katafygiotis, C. Papadimitriou, N.C. Mickleborough, Optimal sensor placement methodology for identification with unmeasured excitation, *Journal of Dynamic Systems, Measurement, and Control*, 23(2011) 677-686.
- [24] E.B. Flynn, M.D. Todd, A Bayesian approach to optimal sensor placement for structural health monitoring with application to active sensing, *Mechanical Systems and Signal Processing*, 24(2010) 891-903.
- [25] J. Von Neumann, O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ, USA, 1953.
- [26] E. Reynders, System identification methods for (operational) modal analysis: Review and Comparison, *Archives of Computational Methods in Engineering*, 19(1)(2012) 51-124.
- [27] Y.Q. Ni, Y. Xia, W.Y. Liao and J.M. Ko, Technology innovation in developing the structural health monitoring system for Guangzhou New TV Tower, *Structural Control and Health Monitoring*, 16(1)(2009) 73-98.
- [28] A. DasGupta, W.J. Studden, Robust Bayesian experimental designs in normal linear models, *The Annals of Statistics*, 19(1991) 1244-1256.
- [29] B. Toman, J.L. Gastwirth, Robust Bayesian experimental design and estimation for analysis of variance models using a class of normal mixtures, *Journal of Statistical Planning and Inference*, 35(3)(1993) 383-398.
- [30] J.A. Thomas, T.M. Cover, *Elements of Information Theory*, 2nd ed., John Wiley & Sons. Inc., Hoboken, NJ, USA, 2006.
- [31] J.M. Bernardo, Expected information as expected utility, *The Annals of Statistics*, 7(1979) 686-690.
- [32] G.C. Tiao, B. Afonja, Some Bayesian considerations of the choice of design for ranking, selection and estimation, *Annals of the Institute of Statistical Mathematics*, 28(1)(1976) 167-185.
- [33] M.L. Eaton, G. Alessandra, S. Paola, A predictive approach to the Bayesian design problem with application to normal regression models, *Biometrika*, 83(1)(1996) 111-125.
- [34] C. Marlings, M. Pozzi, Sensor network optimization using Bayesian networks, decision graphs, and value of information, , in: *Proceeding of 12th International Conference on Applications of Statistics and Probability in Civil Engineering*, Vancouver, Canada, 2015.
- [35] P.H. Garthwaite, J.B. Kadane, O. Anthony, Statistical methods for eliciting probability distributions, *Journal of the American Statistical Association*, 100(2005) 680-701.
- [36] S.H. Crandall, W. D. Mark, *Random Vibration in Mechanical Systems*, Academic Press, New York, USA, 1963.
- [37] A. Der Kiureghian, Structural response to stationary excitation, *Journal of the Engineering Mechanics Division*, 106(1980) 1195-1213.
- [38] R.S. Harichandran, Random vibration under propagating excitation: closed-form solutions, *Journal of Engineering Mechanics*, 118(3)(1992) 575–586.
- [39] T.F. Chan, J.A. Olkin, Circulant preconditioners for Toeplitz-block matrices, *Numerical Algorithms*, 6(1994) 89-101.
- [40] C.F. Van Loan, The ubiquitous Kronecker product, *Journal of Computational and Applied Mathematics*, 123(1-2)(2000) 85-100.
- [41] W.-H. Steeb, T.K. Shi, *Matrix Calculus and Kronecker Product With Applications and C++ Programs*, World Scientific, Singapore, 1997.
- [42] M.G. Genton, Separable approximations of space-time covariance matrices, *Environmetrics*, 18(2007) 681-695.
- [43] Load Code for the Design of Building Structures, China National Standard, GB 50009-2012, China Architecture and Building Press, Beijing, 2012.
- [44] T. Yi, H. Li, M. Gu, Optimal sensor placement for structural health monitoring based on multiple optimization strategies, *The Structural Design of Tall and Special Buildings*, 20(7)(2011) 881-900.
- [45] T. Yi, H. Li, X. Zhang, Sensor placement on Canton Tower for health monitoring using asynchronous-climb monkey algorithm, *Smart Materials and Structures*, 21(2012) 1-12.
- [46] S. Adhikari, Wishart random matrices in probabilistic structural mechanics, *Journal of Engineering Mechanics*, 134(2008) 1029-1044.

Appendix A

In this appendix, we derive the expression for the utility function when employing the expected K-L divergence as described in Section 3.2.3.

Following the notations in Section 2, the following identities hold:

$$\mathbf{Y}_{rs} = (\mathbf{I}_N \otimes \mathbf{S})\mathbf{Y}_r; \mathbf{X}_s = (\mathbf{I}_N \otimes \mathbf{S})\mathbf{X}; \mathbf{W}_s = (\mathbf{I}_N \otimes \mathbf{S})\mathbf{W}; \mathbf{\Sigma}_s = \mathbf{S}\mathbf{\Sigma}\mathbf{S}^T \quad (\text{A.1})$$

Similar to Eqn. (7), we obtain the posterior expectation and covariance of $\boldsymbol{\theta}_s$ given $\mathbf{Y}_r, \mathbf{X}, \boldsymbol{\Gamma}$ and $\mathbf{\Sigma}$ as

$$\begin{aligned} E[\boldsymbol{\theta}_s | \mathbf{Y}_r, \mathbf{X}, \boldsymbol{\Gamma}, \mathbf{\Sigma}] &= \boldsymbol{\Gamma} \mathbf{X}^T [\mathbf{X} \boldsymbol{\Gamma}_s \mathbf{X}^T + \mathbf{I}_N \otimes \mathbf{\Sigma}]^{-1} \mathbf{Y}_r = [\boldsymbol{\Gamma}_s^{-1} + \mathbf{X}^T (\mathbf{I}_N \otimes \mathbf{\Sigma}^{-1}) \mathbf{X}] \mathbf{X}^T (\mathbf{I}_N \otimes \mathbf{\Sigma}^{-1}) \mathbf{Y}_r \quad (\text{A.2}) \\ \text{Cov}[\boldsymbol{\theta}_s | \mathbf{Y}_r, \mathbf{X}, \boldsymbol{\Gamma}, \mathbf{\Sigma}] &= \boldsymbol{\Gamma} - \boldsymbol{\Gamma} \mathbf{X}^T [\mathbf{X} \boldsymbol{\Gamma} \mathbf{X}^T + \mathbf{I}_N \otimes \mathbf{\Sigma}]^{-1} \mathbf{X} \boldsymbol{\Gamma} = [\boldsymbol{\Gamma}^{-1} + \mathbf{X}^T (\mathbf{I}_N \otimes \mathbf{\Sigma}^{-1}) \mathbf{X}]^{-1} \quad (\text{A.3}) \end{aligned}$$

Since posterior distributions of $\boldsymbol{\theta}_s$ for both cases follow multivariate normal distribution with means and covariances as given in Eqn. (7) and the above, the K-L divergence between them is given by (see [30]):

$$KL[\boldsymbol{\theta}_s^{(s)} || \boldsymbol{\theta}_s] = \frac{1}{2} \left[\ln \frac{\det \boldsymbol{\Sigma}_{\boldsymbol{\theta}}}{\det \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(s)}} + \text{tr} \left[\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(s)} \left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(s)-1} \right) \right] + (\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(s)} - \boldsymbol{\mu}_{\boldsymbol{\theta}})^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(s)} - \boldsymbol{\mu}_{\boldsymbol{\theta}}) \right] \quad (\text{A.4})$$

When inserting the preceding equation into Eqn. (13), the first two terms do not involve $\mathbf{Y}_r$; therefore, we only consider the third term, and its expectation with respect to the conditional distribution of $\mathbf{Y}_r$ for given $\mathbf{X}, \boldsymbol{\Gamma}$ and $\mathbf{\Sigma}$ and derive it as follows:

$$\begin{aligned} E[(\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(s)} - \boldsymbol{\mu}_{\boldsymbol{\theta}})^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(s)} - \boldsymbol{\mu}_{\boldsymbol{\theta}})] &= \text{tr} \left\{ E[(\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(s)} - \boldsymbol{\mu}_{\boldsymbol{\theta}})(\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(s)} - \boldsymbol{\mu}_{\boldsymbol{\theta}})^T] \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \right\} \\ &= \text{tr} \left\{ E[\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(s)} \boldsymbol{\mu}_{\boldsymbol{\theta}}^{(s)T}] \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \right\} - 2 \text{tr} \left\{ E[\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(s)} \boldsymbol{\mu}_{\boldsymbol{\theta}}^T] \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \right\} + \text{tr} \{ E[\boldsymbol{\mu}_{\boldsymbol{\theta}} \boldsymbol{\mu}_{\boldsymbol{\theta}}^T] \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \} = \text{tr}(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(s)} \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}) - m \quad (\text{A.5}) \end{aligned}$$

in which,

$$\begin{aligned} \text{tr} \left\{ E[\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(s)} \boldsymbol{\mu}_{\boldsymbol{\theta}}^{(s)T}] \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \right\} &= \text{tr} \{ \boldsymbol{\Gamma}_s \mathbf{X}_s^T [\mathbf{X}_s \boldsymbol{\Gamma}_s \mathbf{X}_s^T + \mathbf{I}_N \otimes \mathbf{\Sigma}_s]^{-1} E[\mathbf{Y}_{rs} \mathbf{Y}_{rs}^T] [\mathbf{X}_s \boldsymbol{\Gamma}_s \mathbf{X}_s^T + \mathbf{I}_N \otimes \mathbf{\Sigma}_s]^{-1} \mathbf{X}_s \boldsymbol{\Gamma}_s \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \} \\ &= \text{tr} \{ \boldsymbol{\Gamma}_s \mathbf{X}_s^T [\mathbf{X}_s \boldsymbol{\Gamma}_s \mathbf{X}_s^T + \mathbf{I}_N \otimes \mathbf{\Sigma}_s]^{-1} \mathbf{X}_s \boldsymbol{\Gamma}_s \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \} = \text{tr}[(\boldsymbol{\Gamma}_s - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(s)}) \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}] \quad (\text{A.6}) \end{aligned}$$

$$\begin{aligned} \text{tr} \left\{ E[\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(s)} \boldsymbol{\mu}_{\boldsymbol{\theta}}^T] \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \right\} &= \text{tr} \{ \boldsymbol{\Gamma}_s \mathbf{X}_s^T [\mathbf{X}_s \boldsymbol{\Gamma}_s \mathbf{X}_s^T + \mathbf{I}_N \otimes \mathbf{\Sigma}_s]^{-1} E[\mathbf{Y}_{rs} \mathbf{Y}_r^T] [\mathbf{X} \boldsymbol{\Gamma}_s \mathbf{X}^T + \mathbf{I}_N \otimes \mathbf{\Sigma}]^{-1} \mathbf{X} \boldsymbol{\Gamma}_s \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \} \\ &= \text{tr} \{ \boldsymbol{\Gamma}_s \mathbf{X}_s^T [\mathbf{X}_s \boldsymbol{\Gamma}_s \mathbf{X}_s^T + \mathbf{I}_N \otimes \mathbf{\Sigma}_s]^{-1} (\mathbf{I}_N \otimes \mathbf{S}) E[\mathbf{Y}_r \mathbf{Y}_r^T] [\mathbf{X} \boldsymbol{\Gamma}_s \mathbf{X}^T + \mathbf{I}_N \otimes \mathbf{\Sigma}]^{-1} \mathbf{X} \boldsymbol{\Gamma}_s \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \} \\ &= \text{tr} \{ \boldsymbol{\Gamma}_s \mathbf{X}_s^T [\mathbf{X}_s \boldsymbol{\Gamma}_s \mathbf{X}_s^T + \mathbf{I}_N \otimes \mathbf{\Sigma}_s]^{-1} (\mathbf{I}_N \otimes \mathbf{S}) \mathbf{X} \boldsymbol{\Gamma}_s \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \} \\ &= \text{tr} \{ \boldsymbol{\Gamma}_s \mathbf{X}_s^T [\mathbf{X}_s \boldsymbol{\Gamma}_s \mathbf{X}_s^T + \mathbf{I}_N \otimes \mathbf{\Sigma}_s]^{-1} \mathbf{X}_s \boldsymbol{\Gamma}_s \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \} = \text{tr}[(\boldsymbol{\Gamma}_s - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(s)}) \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}] \quad (\text{A.7}) \end{aligned}$$

$$\begin{aligned} \text{tr} \{ E[\boldsymbol{\mu}_{\boldsymbol{\theta}} \boldsymbol{\mu}_{\boldsymbol{\theta}}^T] \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \} &= \text{tr} \{ \boldsymbol{\Gamma}_s \mathbf{X}^T [\mathbf{X} \boldsymbol{\Gamma}_s \mathbf{X}^T + \mathbf{I}_N \otimes \mathbf{\Sigma}]^{-1} E[\mathbf{Y}_r \mathbf{Y}_r^T] [\mathbf{X} \boldsymbol{\Gamma}_s \mathbf{X}^T + \mathbf{I}_N \otimes \mathbf{\Sigma}]^{-1} \mathbf{X} \boldsymbol{\Gamma}_s \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \} \\ &= \text{tr} \{ \boldsymbol{\Gamma}_s \mathbf{X}^T [\mathbf{X} \boldsymbol{\Gamma}_s \mathbf{X}^T + \mathbf{I}_N \otimes \mathbf{\Sigma}]^{-1} \mathbf{X} \boldsymbol{\Gamma}_s \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \} = \text{tr}[(\boldsymbol{\Gamma}_s - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}] \quad (\text{A.8}) \end{aligned}$$

Substituting all these expressions into Eqn. (13) gives

$$\begin{aligned} U_3(s) &\propto \iiint \left[\frac{1}{2} \ln \det \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(s)} - \text{tr}(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(s)} \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}) \right] p(\mathbf{X}) p(\boldsymbol{\Gamma}) p(\mathbf{\Sigma}) d\mathbf{X} d\boldsymbol{\Gamma} d\mathbf{\Sigma} \\ &= - \iiint \left\{ \frac{1}{2} \ln \det [\boldsymbol{\Gamma}_s^{-1} + \mathbf{I}_N \otimes \mathbf{F}_s] + \text{tr} \{ [\boldsymbol{\Gamma}_s^{-1} + \mathbf{I}_N \otimes \mathbf{F}_s]^{-1} [\boldsymbol{\Gamma}_s^{-1} + \mathbf{I}_N \otimes \mathbf{F}] \} \right\} p(\mathbf{X}) p(\boldsymbol{\Gamma}) p(\mathbf{\Sigma}) d\mathbf{X} d\boldsymbol{\Gamma} d\mathbf{\Sigma} \quad (\text{A.9}) \end{aligned}$$

Appendix B

Under the input PSD in Eqn.(15), the covariances of the i th and the j th modal responses can be calculated by inverse Fourier transform of the corresponding cross-PSD functions, i.e.,

$$\Gamma_{q_i^{(e)} q_j^{(f)}}(\tau) = (-1)^f I^{e+f} \sum_{k=1}^{N_f} \int_{-\omega_{c,k}}^{\omega_{c,k}} \gamma_{ik} \gamma_{jk} \Phi_{0,k} H_i(\omega) H_j^*(\omega) \omega^{e+f} e^{I\omega\tau} d\omega; e, f = \{0,1,2,3,4\} \quad (B.1)$$

where

$$H_i(\omega) = \frac{1}{\omega_i^2 - \omega^2 + 2I\xi_i \omega_i \omega}; H_j^*(\omega) = \frac{1}{\omega_j^2 - \omega^2 - 2I\xi_j \omega_j \omega} \quad (B.2)$$

In order to evaluate this integral, we first make the partial fraction decomposition of $Q(\omega) = H_i(\omega) H_j^*(\omega)$ based on its poles and residues. For $\xi_i < 1$, this function has four poles so that using partial fractions decomposition one can write

$$Q(\omega) = \sum_{l=1}^4 \frac{Res(\Omega_l)}{\omega - \Omega_l} \quad (B.3)$$

where $Res(\Omega_l)$ is the residue corresponding to pole Ω_l and has the following form:

$$\begin{cases} \Omega_1 = \omega_i \sqrt{1 - \xi_i^2} + I\xi_i \omega_i; & Res(\Omega_1) = \frac{1}{(\Omega_1 - \Omega_2)(\Omega_1 - \Omega_3)(\Omega_1 - \Omega_4)} \\ \Omega_2 = -\omega_i \sqrt{1 - \xi_i^2} + I\xi_i \omega_i; & Res(\Omega_2) = \frac{1}{(\Omega_2 - \Omega_1)(\Omega_2 - \Omega_3)(\Omega_2 - \Omega_4)} \\ \Omega_3 = \omega_j \sqrt{1 - \xi_j^2} - I\xi_j \omega_j; & Res(\Omega_3) = \frac{1}{(\Omega_3 - \Omega_1)(\Omega_3 - \Omega_2)(\Omega_3 - \Omega_4)} \\ \Omega_4 = -\omega_j \sqrt{1 - \xi_j^2} - I\xi_j \omega_j; & Res(\Omega_4) = \frac{1}{(\Omega_4 - \Omega_1)(\Omega_4 - \Omega_2)(\Omega_4 - \Omega_3)} \end{cases} \quad (B.4)$$

Then, substituting the partial fraction decomposition of $Q(\omega)$ gives Eqn. (16), and the needed integral has a closed form solution as follows:

If $\tau = 0$,

$$\int_{-\omega_{c,k}}^{\omega_{c,k}} \frac{\omega^n e^{I\omega\tau}}{\omega - \Omega_l} d\omega = \int_{-\omega_{c,k}}^{\omega_{c,k}} \frac{\omega^n}{\omega - \Omega_l} d\omega = \Omega_l^n \ln \frac{\omega_{c,k} - \Omega_l}{-\omega_{c,k} - \Omega_l} + \sum_{m=1}^n \frac{\Omega_l^{n-m}}{m} [\omega_{c,k}^m - (-1)^m \omega_{c,k}^m] \quad (B.5)$$

If $\tau \neq 0$,

$$\int_{-\omega_{c,k}}^{\omega_{c,k}} \frac{\omega^n e^{I\omega\tau}}{\omega - \Omega_l} d\omega = \Omega_l^n e^{I\Omega_l\tau} \{Ei[I\tau(\omega_{c,k} - \Omega_l)] - Ei[-I\tau(\omega_{c,k} + \Omega_l)]\} + \sum_{m=1}^n \Omega_l^{m-1} \int_{-\omega_{c,k}}^{\omega_{c,k}} \omega^{n-m} e^{I\omega\tau} d\omega \quad (B.6)$$

where

$$\int_{-\omega_{c,k}}^{\omega_{c,k}} \omega^{n-m} e^{I\omega\tau} d\omega = -I \sum_{p=0}^{n-m} \frac{(n-m)!}{(n-m-p)!} \frac{e^{I p \pi/2}}{\tau^{p+1}} [\omega_{c,k}^{n-m-p} e^{I\tau\omega_{c,k}} - (-1)^{n-m-p} \omega_{c,k}^{n-m-p} e^{-I\tau\omega_{c,k}}] \quad (B.7)$$

In the above, $\ln(z)$ denotes the multivalued complex logarithmic function and $Ei(z)$ is the exponential integral function

$$Ei(z) = \int_{-\infty}^z \frac{e^t}{t} dt \quad (\text{B.8})$$

for any complex z not equal to zero and not on the positive real axis. Note that $Ei(z)$ is a multivalued function whose Cauchy principal value jumps by $2\pi i$ when the negative real axis is crossed. Accordingly, $2\pi i$ must be deducted from the difference between the Cauchy principal value of $Ei[I\tau(\omega_{c,k} - \Omega_l)]$ and $Ei[-I\tau(\omega_{c,k} + \Omega_l)]$, if $-I\tau(\omega_{c,k} + \Omega_l)$ locates in the 3th quadrant and $I\tau(\omega_{c,k} - \Omega_l)$ in the 2nd quadrant. It is easy to show that $I\tau(\omega_{c,k} - \Omega_l)$ and $-I\tau(\omega_c + \Omega_l)$ share the same real part, and that the imaginary part of $-I\tau(\omega_c + \Omega_l)$ is always negative while the imaginary part of $I\tau(\omega_{c,k} - \Omega_l)$ is positive. Therefore, this adjustment needs to be made for two of the four poles.

Appendix C

Instead of approximating the covariance matrix, we first try to find the nearest Kronecker product of the correlation matrix, because it usually gives better numerical stability. Assume we have $\mathbf{\Gamma}_s = \mathbf{D}\mathbf{R}\mathbf{D}$, in which $\mathbf{\Gamma}_s \in \mathbb{R}^{k \times k}$ is the covariance matrix, $\mathbf{R} \in \mathbb{R}^{k \times k}$ is the correlation matrix, and $\mathbf{D} \in \mathbb{R}^{k \times k}$ is a diagonal matrix consisting of standard deviations. Recalling that $\mathbf{\Gamma}_s$ is a block Toeplitz matrix, we can express $\mathbf{D} = \mathbf{I}_N \otimes \mathbf{D}_0$ where $\mathbf{D}_0 \in \mathbb{R}^{m \times m}$ is composed of the standard deviations of modal responses. Suppose we have the nearest kronecker product approximation $\mathbf{R} \cong \mathbf{A} \otimes \mathbf{B}$, where $\mathbf{A} \in \mathbb{R}^{k_1 \times k_1}$, $\mathbf{B} \in \mathbb{R}^{k_2 \times k_2}$, and $k = k_1 k_2$, then

$$\mathbf{\Gamma}_s = (\mathbf{I}_N \otimes \mathbf{D}_0)(\mathbf{A} \otimes \mathbf{B})(\mathbf{I}_N \otimes \mathbf{D}_0) = \mathbf{A} \otimes [(\mathbf{I}_{N/k_1} \otimes \mathbf{D}_0)\mathbf{B}(\mathbf{I}_{N/k_1} \otimes \mathbf{D}_0)] \quad (\text{C.1})$$

Perform the following Cholesky decomposition:

$$\mathbf{A} = \mathbf{L}_A \mathbf{L}_A^T; [(\mathbf{I}_{N/k_1} \otimes \mathbf{D}_0)\mathbf{B}(\mathbf{I}_{N/k_1} \otimes \mathbf{D}_0)] = \mathbf{L}_B \mathbf{L}_B^T \quad (\text{C.2})$$

Then we have the Cholesky decomposition of the covariance matrix: $\mathbf{\Gamma}_s = (\mathbf{L}_A \otimes \mathbf{L}_B)(\mathbf{L}_A \otimes \mathbf{L}_B)^T$

The posterior covariance is computed as follows:

$$\begin{aligned} [\mathbf{\Gamma}_s^{-1} + \mathbf{I}_N \otimes \mathbf{F}_s]^{-1} &= [(\mathbf{L}_A \otimes \mathbf{L}_B)^{-T}(\mathbf{L}_A \otimes \mathbf{L}_B)^{-1} + \mathbf{I}_N \otimes \mathbf{F}_s]^{-1} \\ &= (\mathbf{L}_A \otimes \mathbf{L}_B)[\mathbf{I}_k + (\mathbf{L}_A \otimes \mathbf{L}_B)^T(\mathbf{I}_N \otimes \mathbf{F}_s)(\mathbf{L}_A \otimes \mathbf{L}_B)]^{-1}(\mathbf{L}_A \otimes \mathbf{L}_B)^T \\ &= (\mathbf{L}_A \otimes \mathbf{L}_B)[\mathbf{I}_k + (\mathbf{L}_A^T \mathbf{L}_A) \otimes [\mathbf{L}_B^T(\mathbf{I}_{N/k_1} \otimes \mathbf{F}_s)\mathbf{L}_B]]^{-1}(\mathbf{L}_A \otimes \mathbf{L}_B)^T \end{aligned} \quad (\text{C.3})$$

Now, perform the eigenvalue decompositions

$$\mathbf{L}_A^T \mathbf{L}_A = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T; \text{ and } \mathbf{L}_B^T (\mathbf{I}_{N/k_1} \otimes \mathbf{F}_s) \mathbf{L}_B = \mathbf{V} \mathbf{\Psi} \mathbf{V}^T \quad (\text{C.4})$$

in which $\mathbf{U} \in \mathbb{R}^{k_1 \times k_1}$ and $\mathbf{V} \in \mathbb{R}^{k_2 \times k_2}$ are orthogonal matrices, $\mathbf{\Lambda} \in \mathbb{R}^{k_1 \times k_1}$ and $\mathbf{\Psi} \in \mathbb{R}^{k_2 \times k_2}$ are eigenvalue diagonal matrices. Substituting these decompositions into the above equation gives

$$\begin{aligned} [\mathbf{\Gamma}^{-1} + \mathbf{I}_N \otimes \mathbf{F}]^{-1} &= (\mathbf{L}_A \otimes \mathbf{L}_B)[\mathbf{I}_k + (\mathbf{U} \otimes \mathbf{V})(\mathbf{\Lambda} \otimes \mathbf{\Psi})(\mathbf{U}^T \otimes \mathbf{V}^T)]^{-1}(\mathbf{L}_A \otimes \mathbf{L}_B)^T \\ &= (\mathbf{L}_A \otimes \mathbf{L}_B)(\mathbf{U} \otimes \mathbf{V})[(\mathbf{U}^T \otimes \mathbf{V}^T)(\mathbf{U} \otimes \mathbf{V}) + \mathbf{\Lambda} \otimes \mathbf{\Psi}]^{-1}(\mathbf{U}^T \otimes \mathbf{V}^T)(\mathbf{L}_A \otimes \mathbf{L}_B)^T \end{aligned}$$

$$= (\mathbf{L}_A \mathbf{U} \otimes \mathbf{L}_B \mathbf{V}) [\mathbf{I}_k + \mathbf{\Lambda} \otimes \mathbf{\Psi}]^{-1} (\mathbf{L}_A \mathbf{U} \otimes \mathbf{L}_B \mathbf{V})^T \quad (\text{C.5})$$

It is easy to see that the matrix in the bracket is diagonal so that the calculation of its inverse becomes trivial. Therefore, the utility function $U_2(d)$ can be easily evaluated by

$$\ln \det[\mathbf{\Gamma}_s^{-1} + \mathbf{I}_N \otimes \mathbf{F}_s] = -2 \ln \det(\mathbf{L}_A \otimes \mathbf{L}_B) + \ln \det(\mathbf{I}_k + \mathbf{\Lambda} \otimes \mathbf{\Psi}) \propto \ln \det(\mathbf{I}_k + \mathbf{\Lambda} \otimes \mathbf{\Psi}) \quad (\text{C.6})$$

For the utility function $U_1(d)$, we have

$$\begin{aligned} \text{tr}\{[\mathbf{\Gamma}_s^{-1} + \mathbf{I}_N \otimes \mathbf{F}_s]^{-1}\} &= \text{tr}\{(\mathbf{L}_A \mathbf{U} \otimes \mathbf{L}_B \mathbf{V})^T (\mathbf{L}_A \mathbf{U} \otimes \mathbf{L}_B \mathbf{V}) [\mathbf{I}_k + \mathbf{\Lambda} \otimes \mathbf{\Psi}]^{-1}\} \\ &= \text{tr}\{[(\mathbf{U}^T \mathbf{L}_A^T \mathbf{L}_A \mathbf{U}) \otimes (\mathbf{V}^T \mathbf{L}_B^T \mathbf{L}_B \mathbf{V})] [\mathbf{I}_k + \mathbf{\Lambda} \otimes \mathbf{\Psi}]^{-1}\} \\ &= \text{tr}\{[\mathbf{\Lambda} \otimes (\mathbf{V}^T \mathbf{L}_B^T \mathbf{L}_B \mathbf{V})] [\mathbf{I}_k + \mathbf{\Lambda} \otimes \mathbf{\Psi}]^{-1}\} \end{aligned} \quad (\text{C.7})$$

Since matrices in the brackets are block diagonal and diagonal matrix individually, their multiplication only involves diagonal blocks, which is easily calculated.

Now, let us consider the trace term in utility function $U_3(d)$.

$$\begin{aligned} &\text{tr}\{[\mathbf{\Gamma}_s^{-1} + \mathbf{I}_N \otimes \mathbf{F}_s]^{-1} [\mathbf{\Gamma}_s^{-1} + \mathbf{I}_N \otimes \mathbf{F}]\} \\ &= \text{tr}\{(\mathbf{L}_A \mathbf{U} \otimes \mathbf{L}_B \mathbf{V}) [\mathbf{I}_k + \mathbf{\Lambda} \otimes \mathbf{\Psi}]^{-1} (\mathbf{L}_A \mathbf{U} \otimes \mathbf{L}_B \mathbf{V})^T (\mathbf{L}_A \mathbf{U} \otimes \mathbf{L}_B \mathbf{V}')^{-T} [\mathbf{I}_k + \mathbf{\Lambda} \otimes \mathbf{\Psi}'] (\mathbf{L}_A \mathbf{U} \otimes \mathbf{L}_B \mathbf{V}')^{-1}\} \\ &= \text{tr}\{(\mathbf{I}_{k_1} \otimes \mathbf{V}'^T \mathbf{V}) [\mathbf{I}_k + \mathbf{\Lambda} \otimes \mathbf{\Psi}]^{-1} (\mathbf{I}_{k_1} \otimes \mathbf{V}^T \mathbf{V}') [\mathbf{I}_k + \mathbf{\Lambda} \otimes \mathbf{\Psi}']\} \end{aligned} \quad (\text{C.8})$$

Here, we have employed the eigenvalue decomposition $\mathbf{L}_B^T (\mathbf{I}_{N/k_1} \otimes \mathbf{F}) \mathbf{L}_B = \mathbf{V}' \mathbf{\Psi}' \mathbf{V}'^T$. Similar to the utility function $U_2(d)$, each multiplicator is block diagonal or diagonal. Therefore, we can also efficiently evaluate $U_3(d)$.